

# Big Data and Business Intelligence - course description

| General information |                                                                                    |
|---------------------|------------------------------------------------------------------------------------|
| Course name         | Big Data and Business Intelligence                                                 |
| Course ID           | 11.3-WI-INF-D-BDiAB                                                                |
| Faculty             | <a href="#">Faculty of Computer Science, Electrical Engineering and Automatics</a> |
| Field of study      | Computer Science / Computer Systems Engineering                                    |
| Education profile   | academic                                                                           |
| Level of studies    | Second-cycle studies leading to MSc degree                                         |
| Beginning semester  | summer term 2016/2017                                                              |

| Course information  |                                                                        |
|---------------------|------------------------------------------------------------------------|
| Semester            | 1                                                                      |
| ECTS credits to win | 6                                                                      |
| Course type         | obligatory                                                             |
| Teaching language   | polish                                                                 |
| Author of syllabus  | <ul style="list-style-type: none"><li>dr inż. Mariusz Jacyno</li></ul> |

| Classes forms  |                                |                            |                                |                            |                    |
|----------------|--------------------------------|----------------------------|--------------------------------|----------------------------|--------------------|
| The class form | Hours per semester (full-time) | Hours per week (full-time) | Hours per semester (part-time) | Hours per week (part-time) | Form of assignment |
| Lecture        | 30                             | 2                          | 18                             | 1,2                        | Credit with grade  |
| Laboratory     | 30                             | 2                          | 18                             | 1,2                        | Credit with grade  |

## Aim of the course

Zapoznanie studentów z pojęciem analityki biznesowej oraz jej zastosowaniem do analizy dużych zbiorów danych zawartych w mediach społecznościowych, systemach ERP oraz nowoczesnych aplikacjach e-biznesowych.

Nauczenie studentów doboru odpowiednich technik analizy danych w zależności od skali rozpatrywanego problemu oraz rodzaju przeprowadzanej analizy (w czasie rzeczywistym, w trybie batch, przetwarzanie strumieni danych).

Nauczenie studentów pracy z wykorzystaniem nowoczesnych platform analitycznych takich jak: Elastic Search, Apache Hadoop, Apache Spark, Apache Storm oraz SAS

## Prerequisites

Bazy danych,

Znajomość podstaw statystyki

Umiejętność programowania w języku Java

## Scope

Definicja analityki biznesowej. Jej rola i zastosowanie we współczesnych systemach informatycznych.

Przegląd tradycyjnych pojęć i narzędzi analityki biznesowej. Jakość i czyszczenie danych.

Metody redukcji wymiaru: analiza składowych głównych i analiza czynnikowa.

Wnioskowanie statystyczne. Analiza regresji, współzależności i wariancji z punktu widzenia data mining.

Regresja logistyczna. Naiwna estymacja bayesowska i sieci bayesowskie. Analiza i prognozowanie szeregów czasowych.

Analiza danych nieustrukturyzowanych: analiza sentymentu, tworzenie ontologii, kategoryzacja treści, text mining.

Analiza struktury sieci WWW: wyszukiwanie informacji tekstowych i wyszukiwanie w Internecie; ranking oparty o strukturę połączeń.

Analiza użytkowania sieci WWW: wstępne przetwarzanie danych; eksploracyjna analiza użytkowania sieci; grupowanie, reguły asocjacyjne i klasyfikacja.

Nowoczesne narzędzia do zaawansowanej wizualizacji oraz eksploracji danych na przykładzie platformy SAS.

Fenomen Big Data, jego charakterystyka oraz wpływ na istniejące rozwiązania analityczne.

Analityka biznesowa na dużą skalę; nowoczesne rozwiązania wykorzystywane do przesyłania, składowania oraz przetwarzania dużych zbiorów danych.

Architektura nowoczesnych systemów do przetwarzania Big Data na przykładzie platform Elasticsearch, Apache Hadoop, Apache Spark oraz Apache Storm.

Podstawowe techniki uczenia maszynowego wykorzystywane w analityce Big Data.

Analityka danych tekstowych w czasie rzeczywistym z wykorzystaniem platformy ElasticSearch.

## Teaching methods

## Learning outcomes and methods of theirs verification

| Outcome description                                                                                                                                                                                         | Outcome symbols                                                                                                                                                         | Methods of verification                                                   | The class form                                               |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------|--------------------------------------------------------------|
| Student potrafi scharakteryzować cechy nowoczesnych platform do analizy Big Data                                                                                                                            | <ul style="list-style-type: none"><li>• <a href="#">K_W09</a></li><li>• <a href="#">K_W11</a></li></ul>                                                                 | <ul style="list-style-type: none"><li>• a test with score scale</li></ul> | <ul style="list-style-type: none"><li>• Lecture</li></ul>    |
| Student potrafi w praktyczny sposób wykorzystać narzędzia oferowane przez platformę SAS oraz Elastic Search w celu przeprowadzenia zaawansowanej analizy danych oraz ich eksploracji w czasie rzeczywistym. | <ul style="list-style-type: none"><li>• <a href="#">K_W09</a></li><li>• <a href="#">K_W11</a></li></ul>                                                                 | <ul style="list-style-type: none"><li>• a project</li></ul>               | <ul style="list-style-type: none"><li>• Laboratory</li></ul> |
| Student potrafi zdefiniować pojęcie analityki biznesowej oraz podać przykłady jej zastosowania.                                                                                                             | <ul style="list-style-type: none"><li>• <a href="#">K_W12</a></li></ul>                                                                                                 | <ul style="list-style-type: none"><li>• a test with score scale</li></ul> | <ul style="list-style-type: none"><li>• Lecture</li></ul>    |
| Student jest w stanie opisać tradycyjny model analizy danych wykorzystywany w data i text miningu.                                                                                                          | <ul style="list-style-type: none"><li>• <a href="#">K_W09</a></li><li>• <a href="#">K_W12</a></li></ul>                                                                 | <ul style="list-style-type: none"><li>• a test with score scale</li></ul> | <ul style="list-style-type: none"><li>• Lecture</li></ul>    |
| Student rozumie dlaczego dotychczasowe rozwiązania analityczne nie nadają się do analizy Big Data.                                                                                                          | <ul style="list-style-type: none"><li>• <a href="#">K_W12</a></li><li>• <a href="#">K_U01</a></li></ul>                                                                 | <ul style="list-style-type: none"><li>• a test with score scale</li></ul> | <ul style="list-style-type: none"><li>• Lecture</li></ul>    |
| Student potrafi w praktyczny sposób wykorzystać wybraną platformę do przeprowadzenia analizy big data.                                                                                                      | <ul style="list-style-type: none"><li>• <a href="#">K_W09</a></li><li>• <a href="#">K_W11</a></li></ul>                                                                 | <ul style="list-style-type: none"><li>• a project</li></ul>               | <ul style="list-style-type: none"><li>• Laboratory</li></ul> |
| Student zna podstawowe techniki uczenia maszynowego wykorzystywane podczas analizy danych.                                                                                                                  | <ul style="list-style-type: none"><li>• <a href="#">K_W08</a></li><li>• <a href="#">K_W12</a></li><li>• <a href="#">K_U01</a></li><li>• <a href="#">K_U05</a></li></ul> | <ul style="list-style-type: none"><li>• a test with score scale</li></ul> | <ul style="list-style-type: none"><li>• Lecture</li></ul>    |
| Student potrafi wyjaśnić pojęcie Big Data oraz przedstawić główne cechy jakie zawierają zbiory danych określane tą nazwą.                                                                                   | <ul style="list-style-type: none"><li>• <a href="#">K_W12</a></li></ul>                                                                                                 | <ul style="list-style-type: none"><li>• a test with score scale</li></ul> | <ul style="list-style-type: none"><li>• Lecture</li></ul>    |

## Assignment conditions

Wykład - warunkiem zaliczenia jest uzyskanie pozytywnej oceny z egzaminu przeprowadzonego w formie zaproponowanej przez prowadzącego

Laboratorium - warunkiem zaliczenia jest uzyskanie ocen pozytywnych z wszystkich ćwiczeń laboratoryjnych

Składowe oceny końcowej = wykład: 50% + laboratorium: 50%

## Recommended reading

Morzy T.: Eksploracja danych. Metody i algorytmy, PWN, Warszawa, 2013

Larose D.T.: Metody i modele eksploracji danych, PWN, Warszawa, 2008

Markov Z., Larose D.T.: Eksploracja zasobów internetowych, PWN, Warszawa, 2009

White T., Hadoop: The Definitive Guide, 3rd Edition, O'Reilly Media / Yahoo Press, 2012.

George L., HBase: The Definitive Guide, O'Reilly Media, 2011.

## Further reading

Stanton J.M.: Introduction to Data Science, E-book, 2013.

## Notes

Modified by prof. dr hab. inż. Krzysztof Patan (last modification: 22-09-2016 09:13)